University of Montpellier — Year 2024-2025

HAX912X - Genralized Linear Model / Machine Learning

Additional work on Generalized Linear Models

Exercise 1

Consider a n-sample $(y_1, x_1), \ldots, (y_n, x_n)$ of the pair (y, x) where y is a random variable with values in $\{0, 1\}$ and x is a fixed variable also taking the values $\{0, 1\}$. It is assumed that there exists a latent variable y_i^* such that $y_i^*|x_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, 1)$ and

$$y_i = \begin{cases} 1 & \text{si} \quad y_i^* \leq 0 \\ 0 & \text{si} \quad y_i^* > 0 \end{cases} .$$

The distribution function of the standard gaussian distribution is denoted by $\Phi(\cdot)$ ($\Phi(u) = \mathbb{P}(\mathcal{N}(0, 1) \leq u)$).

1 Show that this is a Generalised Linear Model. Following the notations of the course for exponential families, explicit the parameters ϕ, θ, the functions b, c, γ and the regressors to be considered.

2 Has the canonical link been chosen?

3 We have 100 observations (y_i, x_i) described in the following contingency table

	$y = 0$	$y = 1$
$x = 0$	10	30
$x = 1$	20	40

Let $\alpha_0 = \Phi(\beta_0)$ and $\alpha_1 = \Phi(\beta_0 + \beta_1)$ give the log-likelihood as a function of β_0 and β_1.

4 Calculate the maximum likelihood estimator for (α_0, α_1).

Exercise 2

Consider n independent random variables $y_1, \ldots, y_n$ such that $y_i \sim \mathcal{N}\left(\exp(\beta_0 + \beta_1 x_i), \sigma^2\right)$.

1 Show that this is a Generalized Linear Model. Following the notations of the course for exponential families with nuisance parameter, explicit ϕ, θ, the functions γ and b and the regressors to be considered.

2 Show that the canonical link has not been chosen.

3 Assuming in the following that $\sigma^2 = 1$, give the log-likelihood and the likelihood equations. Can we calculate the analytical expressions for the maximum likelihood estimators of β_0 and β_1?